

Started on	Tuesday, 7 January 2020, 3:10 PM
State	Finished
Completed on	Tuesday, 7 January 2020, 3:40 PM
Time taken	29 mins 57 secs
Marks	14.67/15.00
Grade	29.33 out of 30.00 (98%)

Question **1**

Correct

Mark 1.00 out of 1.00

Given the definitions below:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

which of the formulas below computes the accuracy of a binary classifier?

Select one:

- a. $TN / (TN + FP)$
- b. $(TP + TN) / (TP + FP + TN + FN)$
- c. $TP / (TP + FN)$
- d. $TP / (TP + FP)$

Question **2**

Correct

Mark 1.00 out of 1.00

The *information gain* is used to

Select one:

- a. select the class with maximum probability
- b. select the attribute which maximises, for a given test set, the ability to predict the class value
- c. select the attribute which maximises, for a given training set, the ability to predict all the other attribute values
- d. select the attribute which maximises, for a given training set, the ability to predict the class value

Question **3**

Correct

Mark 1.00 out of 1.00

Which of the following is a base hypothesis for a bayesian classifier?

Select one:

- a. The attributes must have negative correlation
- b. The attributes must have zero correlation
- c. The attributes must be statistically independent inside each class
- d. The attributes must be statistically independent

Question **4**

Correct

Mark 1.00 out of
1.00What is the *cross validation***Select one:**

- a. A technique to obtain a good estimation of the performance of a classifier with the training set
- b. A technique to obtain a good estimation of the performance of a classifier when it will be used with data different from the training set
- c. A technique to improve the speed of a classifier
- d. A technique to improve the quality of a classifier

Question **5**

Correct

Mark 1.00 out of
1.00How does *pruning* work when generating frequent itemsets?**Select one:**

- a. If an itemset is not frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated
- b. If an itemset is frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated
- c. If an itemset is not frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated
- d. If an itemset is frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated

Question **6**

Correct

Mark 1.00 out of
1.00

A Decision Tree is...

Select one:

- a. A tree-structured plan of tests on single attributes to obtain the maximum purity of a node
- b. A tree-structured plan of tests on single attributes to forecast the cluster
- c. A tree-structured plan of tests on single attributes to forecast the target
- d. A tree-structured plan of tests on multiple attributes to forecast the target

Question **7**

Correct

Mark 1.00 out of
1.00

Which of the following statements regarding the discovery of association rules is true? (One or more)

Select one or more:

- a. The confidence of a rule can be computed starting from the supports of itemsets
- b. The support of a rule can be computed given the confidence of the rule
- c. The support of an itemset is anti-monotonic with respect to the composition of the itemset
- d. The confidence of an itemset is anti-monotonic with respect to the composition of the itemset

Question **8**

Correct

Mark 1.00 out of 1.00

Which of the following *is not* an objective of feature selection

Select one:

- a. Reduce the effect of noise
- b. Select the features with higher range, which have more influence on the computations
- c. Reduce time and memory complexity of the mining algorithms
- d. Avoid the *curse of dimensionality*

Question **9**

Correct

Mark 1.00 out of 1.00

Which is the main reason for the *normalisation* (also known as "*rescaling*") of numeric attributes?

Select one:

- a. Change the distribution of the numeric attributes, in order to obtain gaussian distributions
- b. Map all the nominal attributes to the same range, in order to prevent the values with higher frequency from having prevailing influence
- c. Map all the numeric attributes to the same range, in order to prevent the attributes with higher range from having prevalent influence
- d. Remove abnormal values

Question **10**

Correct

Mark 1.00 out of 1.00

Which of the statements below is true? (One or more)

Select one or more:

- a. K-means is quite efficient even for large datasets
- b. K-means is very sensitive to the initial assignment of the centers
- c. K-means always stops to a configuration which gives the minimum distortion for the chosen value of the number of clusters.
- d. Sometimes k-means stops to a configuration which does not give the minimum distortion for the chosen value of the number of clusters.

Question **11**

Correct

Mark 1.00 out of 1.00

Which of the following clustering methods is ***not*** based on distances between objects?

Select one:

- a. Expectation Maximization
- b. Hierarchical Agglomerative
- c. DBSCAN
- d. K-Means

Question **12**
Partially correct
Mark 0.67 out of 1.00

Which of the statements below is true? (One or more)

Select one or more:

- ☐ a. Increasing the radius of the neighbourhood can decrease the number of noise points
- ☐ b. DBSCAN always stops to a configuration which gives the optimal number of clusters
- ☐ c. Sometimes DBSCAN stops to a configuration which does not include any cluster
- ☐ d. DBSCAN can give good performance when clusters have concavities

Question **13**
Correct
Mark 1.00 out of 1.00

Which of the following *is not* an objective of feature selection

Select one:

- ☐ a. Avoid the *curse of dimensionality*
- ☐ b. Reduce the effect of noise
- ☐ c. Select the features with higher range, which have more influence on the computations
- ☐ d. Reduce time and memory complexity of the mining algorithms

Question **14**
Correct
Mark 1.00 out of 1.00

Given the two binary vectors below, which is their similarity according to the Simple Matching Coefficient?

a b c d e f g h i j
1 0 0 0 1 0 1 1 0 1
1 0 1 1 1 0 1 0 1 0

Select one:

- ☐ a. 0.1
- ☐ b. 0.3
- ☐ c. 0.5
- ☐ d. 0.2

Question **15**
Correct
Mark 1.00 out of 1.00

Which of the following *is not* a strength point of *Dbscan* with respect to *K-means*

Select one:

- ☐ a. The *effectiveness*, even in presence of *noise*
- ☐ b. The *robustness* with respect to the number of attributes
- ☐ c. The *effectiveness* even if there are clusters with non-convex shape
- ☐ d. The efficiency even in large datasets

